

## Proposition de post-doctorat à partir d'octobre 2011 :

### **«Construction de métamodèles et analyse de sensibilité pour un code de dispersion atmosphérique»**

**Durée du post-doc : 12 à 24 mois**

**Localisation : Laboratoire de Sûreté et Maîtrise des Risques (LSMR), CEA de Cadarache**

**Encadrantes CEA: Amandine MARREL (LSMR), Nadia PEROT (LSMR)**

**Encadrante INSA/IMT: Béatrice LAURENT**

#### **Description du sujet**

Le CEA a développé un code de dispersion atmosphérique bidimensionnel pour l'évaluation des conséquences d'un rejet de produits radioactifs dans l'environnement appelé CERES-MITRA (C-M). Ce code, basé sur un modèle à bouffées gaussiennes (Hanna et al. [1982], a été développé d'une part, pour les études accidentelles des différents dossiers réglementaires des installations du CEA, d'autre part, pour la gestion d'un accident nucléaire survenant sur un site du CEA. Les entrées de cet outil constituent le scénario de calcul. Elles sont nombreuses et de différents types : variables scalaires (vitesse de dépôt des radionucléides, ...), variables fonctionnelles (chroniques de rejet, chroniques météorologiques), variable qualitative ou catégorielle (choix des formules semi-empiriques de calcul des écarts-types du modèle gaussien telles que proposées par Doury [1976], Briggs [1973] ou Turner [1964]). Le code de calcul C-M fournit en sortie les concentrations de radionucléides dans l'air et au sol à différents instants du scénario.

Bien que le modèle C-M ait été largement étudié, validé et utilisé, sa sensibilité aux différents paramètres d'entrée n'a jamais été évaluée dans sa globalité (pas de prise en compte de la variabilité temporelle des données météorologiques, interactions entre les variables non étudiées, ...). En effet, des études paramétriques ont déjà été réalisées mais ce code présente la particularité d'avoir des dizaines de paramètres d'entrée de différents types et de produire des sorties spatio-temporelles, ce qui rend les méthodes habituelles d'analyse de sensibilité inadaptées. Un premier objectif serait donc de développer des méthodes permettant de réaliser une analyse de sensibilité, *i.e.* de quantifier l'influence des différents paramètres d'entrée sur la sortie du code.

En pratique, lorsqu'un calcul C-M est réalisé faisant suite à rejet accidentel, les valeurs des nombreux paramètres ne sont pas exactement connues (identifiées seulement à un ordre de grandeur prêt) et sont entachées d'une plus ou moins grande incertitude. Il serait intéressant d'évaluer l'erreur commise par ce manque de connaissance sur les prévisions d'évolution du panache. Concernant les données météorologiques, elles sont souvent disponibles au moment du rejet (capteurs sur le site) mais sont souvent considérées, pour le lancement du calcul par C-M, comme constantes ou faiblement variables dans le temps. Il serait très utile de quantifier l'erreur commise lorsque l'on néglige la variabilité des chroniques météo sur la durée du rejet. Un second objectif serait donc de réaliser la propagation d'incertitudes des différentes variables d'entrée, celles dont la connaissance est partielle et celles dont la variabilité est souvent négligée.

Répondre à ces différentes problématiques nécessite donc de mettre en œuvre des outils statistiques pour réaliser l'analyse de sensibilité du modèle de transport C-M et propager les incertitudes des différents paramètres d'entrée.

## Problématiques et pistes de recherche envisagées

De par la nature des paramètres d'entrées et des variables de sortie considérées, l'analyse de sensibilité et la propagation des incertitudes du modèle C-M fait apparaître différentes sous-problématiques :

- **Modélisation des sources d'incertitudes** : préalablement à toute étude d'analyse de sensibilité ou de propagation d'incertitudes, il est nécessaire de définir et de quantifier l'incertitude sur les différents paramètres d'entrée. Des données bibliographiques existent pour certaines variables comme la vitesse de dépôt qui se déduit du spectre granulométrique du rejet. Concernant les chroniques de rejet et les chroniques météorologiques, elles sont définies sous forme de paliers, eux aussi définis par l'utilisateur. Il s'agit donc d'entrées fonctionnelles mais constantes par morceaux. Pour les chroniques de rejet, il est possible d'identifier par avis d'experts des chroniques de rejet représentatives de différents types d'accidents et de définir des intervalles de variation pour les quantités rejetées aux différents paliers. Concernant les chroniques météo, le problème est plus complexe car il s'agit de définir des réalisations représentatives des conditions météorologiques possibles sur le site et cohérentes dans le temps et entre elles. Pour cela, on dispose de plusieurs années de relevés météo sur le centre et un stage, actuellement en cours au LSMR, a pour objectif l'analyse approfondie de cette base de données. Cela devrait permettre d'identifier les corrélations entre les variables et de générer des réalisations « réalistes » des chroniques météo. **Le post-doctorant pourra s'appuyer sur les résultats obtenus à l'issue du stage.**
- **Analyse de sensibilité avec entrées fonctionnelles** : la définition et l'estimation d'indices de sensibilité pour des entrées fonctionnelles (chroniques temporelles) nécessitent le développement et l'utilisation d'outils spécifiques. Une possibilité serait, par exemple, de les traiter comme des entrées catégorielles, *i.e.* comme un type de scénario (Ruffo et al. [2006]). **Une paramétrisation ou une décomposition sur une base de fonctions orthogonales peut aussi être envisagée en gardant à l'esprit le besoin d'interpréter les indices de sensibilité qui seront estimés.**
- **Analyse de sensibilité avec entrées corrélées** : Certaines variables d'entrée (variables météo en particulier) peuvent être fortement corrélées. **Dans ce cas, la définition d'indices de sensibilité adaptés demeure un problème qu'il faudra d'étudier.**
- **Analyse de sensibilité et propagation d'incertitude pour des sorties spatio-temporelles** : l'analyse de sensibilité globale est basée sur l'estimation des indices de Sobol (Sobol [1993], Saltelli et al. [2000]). Une estimation empirique de ces indices nécessite un grand nombre d'appels au code C-M, ce qui n'est pas réalisable compte-tenu du grand nombre de paramètre et du temps de calcul nécessaire pour chaque simulation (une dizaine de minutes de temps de calcul en configuration maximale). Pour s'affranchir de ce problème, le code est alors remplacé par un méta-modèle construit à partir d'un nombre limité de simulations et requérant un temps d'évaluation négligeable. Un méta-modèle et une méthode d'analyse de sensibilité pour des sorties spatiales ont été développées dans le cadre d'une précédente thèse au CEA de Cadarache (Marrel [2008]). Ces méthodes s'appuient sur une réduction de la dimension via une projection des sorties spatiales sur une base d'ondelettes. Les coefficients « importants » de la projection sur la base d'ondelettes sont ensuite modélisés, en fonction des paramètres d'entrée du code, par des métamodèles du type processus gaussiens (Sacks et al. [1989]). Le stage, en cours au LSMR, a pour objectif d'appliquer cette méthodologie au code C-M et de la comparer à une seconde approche basée sur une décomposition POD (Proper Orthogonal Décomposition, Chatterjee [2000]) et sur une modélisation des coefficients de la POD en fonction des variables d'entrée. **Durant ce stage, seul l'aspect spatial de la sortie sera traité et l'un des objectifs du post-doctorat sera**

**d'étendre ces travaux au cadre spatio-temporel afin de réaliser l'analyse de sensibilité et la propagation des incertitudes du code C-M.**

- **Grand nombre de paramètres d'entrées :** le nombre de paliers considérés pour les chroniques météorologiques et de rejet peut augmenter considérablement le nombre de variables d'entrée. Au-delà d'une quinzaine de variables, il est conseillé de réaliser une première étape de sélection de variables préalablement à la construction d'un métamodèle. Pour cela, un criblage pourra être réalisé afin d'explorer rapidement le comportement des sorties du code en fonction des variables d'entrée. Plusieurs méthodes de criblage sont envisagées : plans sursaturés utilisés par Claeys-Bruno et al. [2007], criblage par groupe de Dean & Lewis [2006] ou encore bifurcations séquentielles de Bettonvil & Kleijnen [1997]. **Le choix de la méthode sera fonction du nombre de variables considéré, du nombre de simulations possible et de la connaissance a priori disponible sur le sens de variation de la sortie en fonction du sens de variation de chaque entrée. Un des objectifs sera d'étendre la méthode de criblage choisie au cas des entrées fonctionnelles.**

### Contexte du sujet

Ce sujet de post-doctorat est financé par le projet COSTA BRAVA de l'Agence Nationale pour la Recherche. Le site web du projet fournit de nombreuses informations : [http://www.math.univ-toulouse.fr/COSTA\\_BRAVA/doku.php?id=index](http://www.math.univ-toulouse.fr/COSTA_BRAVA/doku.php?id=index). Ce projet a pour objectif de développer de nouvelles méthodes combinant approches déterministes et stochastiques pour réaliser l'analyse de sensibilité des codes de calcul simulant des dynamiques spatio-temporelles. Ce post-doc s'inscrit ainsi dans deux des axes de recherche du projet :

- l'axe 1 sur la comparaison et combinaison d'approches stochastiques et déterministes pour l'analyse de sensibilité ;
- l'axe 3 sur les modèles pour les dynamiques spatio-temporelles instationnaires.

Ce projet réunit des chercheurs des milieux académiques (universités, CNRS, INIRA) et industriels (CEA, IFP Energies nouvelles).

### Compétences requises

Thèse en probabilités/statistique, bonne connaissance des logiciels Matlab et R.

### Contact

Amandine MARREL, CEA Cadarache, CEA/DEN/DER/SESI/LSMR, 13108 Saint-Paul-lez-Durance.

**Email :** [amandine.marrel \[at\] cea.fr](mailto:amandine.marrel@cea.fr)

## Références bibliographiques

Briggs G.A. (1973). Diffusion estimation of small emissions. *Atmospheric Turbulence and Diffusion Laboratory Contribution*, 79:83–145.

Chatterjee A. (2000). An introduction to the proper orthogonal decomposition. *Current Science*, 78:808-817.

Claeys-Bruno M., Dobrijevic M., Cela R., Phan-Tan-Lu R., and Sergent M. (2007). Supersaturated design: construction, comparison and interpretation. *In VI Colloquium Chemiometricum Mediterraneum*, Saint Maximin La Sainte Baume, France, september 2007.

Doury A. (1976). Une méthode de calcul pratique et générale pour la prévision des pollutions véhiculées par l'atmosphère. *Rapport CEA-R-4280*.

Dean A. and Lewis S., editors (2006). Screening - Methods for experimentation in industry, drug discovery and genetics. *Springer*.

Hanna S., Briggs G. and Hosker Jr. R. (1982). Handbook on atmospheric diffusion. *Technical Information Center*, US Department of Energy.

Kleijnen J.P.C., Bettonvil B.W.M. (1997). Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research*, 96(1):180-194.

Marrel A. (2008). Mise en œuvre et utilisation du métamodèle processus gaussien pour l'analyse de sensibilité de modèles numériques. *Thèse de l'INSA Toulouse*.

Ruffo, P., Bazzana, L., Consonni, A., Corradi, A., Saltelli, A., and Tarantola, S. (2006). Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analysis techniques. *Reliability Engineering and System Safety*, 91:1155–1162.

Sacks J., Welch W.J., Mitchell T.J. and Wynn H.P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4:409-435 .

Saltelli A., Chan K., and Scott E.M., editors (2000). Sensitivity analysis. *Wiley Series in Probability and Statistics*, Wiley.

Sobol I.M. (1993). Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414.

Turner D.B. (1964). A diffusion model for an urban area. *Journal of Applied Meteorology*, 3(1):83-91.